DOCUMENT RESUME

ED 210 274 TM 810 674

AUTHOR Tobin, Kenneth G.; Capie, William
TITLE Measuring Pupil Engagement.
PUB DATE Apr 81
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Classification; *Classroom Research; Data Collection; Intermediate Grades; Learning Processes; *Measurement Techniques; Process Education; Science Education; *Student Behavior; *Student Characteristics; Teacher Behavior; Teaching Methods; *Time on Task
IDENTIFIERS Generalizability Theory

ABSTRACT

This paper advocates categorization of engagement on the basis of a logical relationship with the outcomes of a study and the use of student attributes that are logically related to engagement and/or achievement as covariables. Results from a study involving nine engagement categories, measures of formal reasoning ability, locus of control and science achievement are used to highlight issues related to student engagement measures. The possibility of increasing the reliability of selected engagement categories through manipulation of student behavior is discussed. Generalizability theory is used to plan reliable data collection procedures and to assess the reliability of engagement measures. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED210274

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

MEASURING PUPIL ENGAGEMENT

Kenneth G Tobin
Center for Science Education
Mount Lawley College
Perth Australia 6050

William Capie
Department of Science Education
University of Georgia
Athens GA 30602

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
K. Tobin
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.

TM 810 674

MEASURING PUPIL ENGAGEMENT

Research in classrooms has recently focussed on the mediating role of pupil behaviors on learning. Several studies have shown pupil time on-task to be significantly related to pupil achievement (e.g. Hecht, 1978; Ladaherne, 1968; Samuels and Turnure, 1974) while others have confirmed the link between pupil time on-task and teacher performance (e.g. Cornbleth and Korth, 1979; Goldstein and Weber, 1979). In most of the research time on-task has been measured dichotomously in terms of on-task behaviors and off-task behaviors. Although significant positive relationships with achievement have been obtained for the time on-task variables, there is intuitive appeal in the notion that certain types of pupil engagement represent appropriate practice for the intended outcomes while other types of engagement do not. Logical analysis of intended outcomes can lead to the identification of pupil engagement categories that constitute appropriate practice. Stallings (1980) recently reported a study of basic reading skills in which pupil engagement was categorized in terms of nine on-task categories and three off-task categories. She reported significant positive relationships for certain on-task categories, zero correlations for others, and in some cases significant negative correlations with the Comprehensive Test of Basic Skills. Tobin and Capie (1981 a, b) also reported that some types of pupil engagement were related to science achievement while others were not. The identification of achievement-related engagement modes and concomitant teaching strategies can directly benefit classroom practice. For example, if replicable outcomes are obtained from engagement research, teachers

can be encouraged to adopt strategies to enhance rates of specific types of pupil engagement.

In the sections that follow, procedures adopted by Tobin and Capie (1981 a, b) and results from their research in middle school science classrooms are used to highlight important issues related to measurement of pupil engagement.

Categorizing Pupil Engagement

The category system that is most appropriate for a study is dependent on the research questions that prompted the study. The procedure for identifying appropriate categories, however, is generalizable. Tobin and Capie developed an observation instrument to measure pupil engagement on a set of categories related to process skill achievement. Although process skills are usually taught in science programs, they are generic intellectual skills that are applicable to a wide range of problem solving contexts.

Thus, pupil behaviors were identified that were logically involved in planning and conducting an investigation. Additional behaviors that frequently occurred in science classrooms were also identified and defined so that every behavior to occur in the classroom could be rated. Engagement was categorized in terms of seven overt on-task categories, one covert on-task category, and one off-task category.

Sampling Pupil Behavior

Because of the impossibility of measuring everything that happens in a classroom, pupil behavior needs to be sampled. As a consequence, data gathering procedures need to be used to assure that dependable engagement measures are obtained. An indication of the dependability of mean pupil engagement rates for differentiating pupils or classes can be obtained using generalizability theory. When individual pupils from different classes are involved in a study, however, a statistical problem exists in that pupils are nested within classes. As a consequence, the variance attributable to pupils is confounded with the variance attributable to the pupil by class interaction. To overcome this problem, variance components can be estimated separately for each class. Mean component estimates over all classes can then be used as a basis for estimating a generalizability coefficient. If pupil engagement is measured by one rater, a three-facet hierarchical design is applicable for each class with sampling occasion nested within lesson which is nested within pupils. If the data are to be used to differentiate classes, however, a four-facet hierarchical design is applicable with pupils nested within classes. For each of the above examples, unbiased variance component estimates can be calculated from analysis of variance expected mean square terms. Algebraic expressions applicable to three- and four-facet designs are contained in Table 1.

Insert Table 1 about here

A common application of generalizability theory in classroom research is in studies (G-studies) designed to obtain information on optimal data collecting procedures to be adopted in subsequent research (D-studies). In most D-study research a generalizability coefficient greater than .8 is considered necessary in order to minimize the probability of committing type II errors. By estimating variance components from data collected in a G-study, data collection procedures can be simulated so that variation in estimates of the generalizability coefficient can be observed. For example, in a three-facet design, variance components estimated from a G-study could be used as a basis for calculating generalizability coefficients for a range of lessons and sampling occasions. An assumption underlying the use of generalizability theory in this way is that variance components estimated from the G-study are stable, unbiased estimates of population values. To assure tolerable stability, G-studies should involve as many levels on each facet as possible. Complete discussion of variance component stability is contained in Smith (1978, 1980) and Tobin and Capie (1981 c).

The data collection procedures used by Tobin and Capie (1981 a, b) required pupil engagement to be measured in a sequence of eight science lessons presented over a two-week period of time. Each lesson was planned to promote the acquisition of the same set of objectives. Twelve pupils were randomly selected from thirteen middle school classes. Engagement measures were obtained for each pupil for one lesson of every two, thus, six pupils were observed during each lesson. A time sampling procedure was used in which behavior was

observed for each pupil for four seconds during each 30 second interval. If more than one codable behavior occurred in a four second interval, the predominant behavior to occur in that interval was recorded. The order in which pupils were observed in each lesson was randomly determined. In addition to pupil engagement, the predominant teacher behavior to occur in each 30 second interval was also recorded for each lesson. At a later time the recorded teaching behavior in each time interval can be used to categorize covert pupil engagement. For example, if a pupil was attending to instruction while the teacher was constructing a hypothesis from a dataset, the attending category initially assigned to the pupil may be subsequently changed to reflect hypothesizing.

The algebraic expression for calculating the generalizability coefficient for the three facet example is:

$$\rho^2 = \frac{\hat{\sigma}^2_S}{\hat{\sigma}^2_S + \frac{1}{L}\hat{\sigma}^2_{L(S)} + \frac{1}{PL}\hat{\sigma}^2_{RES}}$$

where $\hat{\sigma}^2_S$ is the estimated variance component attributable to pupils;

$\hat{\sigma}^2_{L(S)}$ is the estimated variance component attributable to lessons nested within pupils;

L is the number of lessons;

P is the number of sampling occasions;

$\hat{\sigma}^2_{RES}$ is the estimated residual variance component.

Using this expression, a matrix similar to that provided in Table 2 can be constructed for each engagement mode. The results in Table 2 show how the magnitude of the generalizability coefficient varies as the number of lessons and the number of sampling occasions per lesson are varied. The generalizability coefficients reported by Tobin and Capie (1981 a, b) indicated that dependable data could be obtained from 4 lessons and 40 sampling occasions. However, a preferable data collection system than that used by Tobin and Capie (1981 a, b) would be to measure engagement for each of the twelve pupils in every second lesson. The main advantage of this procedure is that statistical problems attributable to differences between consecutive lessons are avoided.

Insert Table 2 about here

The results in Table 2 suggest that dependable data could also be obtained from six lessons and ten sampling occasions. Since competing designs allow for the collection of dependable data, decision making can take account of other criteria such as convenience to personnel and budget implications.

Tobin and Capie (1981 a, b) estimated generalizability coefficients separately for each engagement mode. Although identical data collection procedures were used for each engagement mode, the generalizability coefficients ranged from .92 for the off-task category to .25 for quantifying. Coefficients of a satisfactory magnitude were only obtained for attending, collecting, and off-task

behavior. The measures obtained for categories that infrequently occurred were not dependable for differentiating pupils. Researchers are provided with several alternatives for increasing the magnitude of the generalizability coefficient: the number of lessons observed could be increased; the number of behaviors sampled per lesson could be increased; or the frequency of occurrence of the behavior to be sampled could be increased. Each of these possibilities is discussed below.

As an alternative to varying the data collection procedures, the occurrence of the actual behaviors of concern may be increased. The most compelling reason for opting for this alternative is that many of the engagement rates appear to have been too low to have had a substantial impact on achievement. One means of achieving an increase would be to incorporate covert engagement as well as overt engagement into the definition of each engagement mode. A second approach is to increase the opportunities for overt and covert engagement on a particular mode by modifying the lesson plans provided to teachers in a study. For example, if teachers were provided with lesson plans that placed an emphasis on overt pupil generalizing, higher rates of generalizing behavior might be anticipated in classes taught by these teachers compared with classes where a greater emphasis on data collecting was incorporated into lesson plans.

In several instances correlations reported by Tobin and Capie (1981 a, b) of engagement modes with achievement were undoubtedly truncated by low dependability. Adequate tests of hypotheses

necessitate the use of dependable measures. As a consequence, care should be taken to use techniques to assure that adequate engagement levels and variability are obtained for each engagement mode. A procedure for assisting in this regard is to use modified lesson plans together with feedback to teachers. In this way pupil engagement rates can be manipulated.

Pupil Engagement and Achievement

The decision to categorize engagement was justified from results reported by Tobin and Capie (1981 a, b). The highest correlates of science achievement were attending ($r = .31$; $p < .001$) and generalizing ($r = .36$, $p < .0001$). Students who attended to instruction tended to achieve at a higher level than those who did not. Similarly, those pupils who exhibited higher rates of generalizing behavior also tended to attain higher achievement scores. Thus two engagement categories were significantly related to achievement while others were not. The validation of the attending mode raises a question on the relative importance of overt and covert pupil behavior. Since the frequency of occurrence of several of the engagement modes (e.g. planning) was close to zero, the principal means of engaging in those categories was covert through the attending mode. By rating the predominant teacher behavior during the same time interval in which pupils were rated, the behaviors to which pupils were attending was recorded. Thus for each engagement mode a covert and overt measure was available for each pupil. Although a comparison on the relative merits of overt and covert engagement is a

possibility, the combined data may be more valid than either data considered separately.

When total engagement was considered, Tobin and Capie (1981 b) reported three engagement modes to be related to pupil achievement. Comprehending ($r = .24$, $p < .001$), planning ($r = .24$, $p < .001$), and generalizing ($r = .30$, $p < .0001$) were each significantly related to summative achievement and retention. These results supported the viewpoint that appropriate practice on the instructional objectives results in pupil achievement increases.

Pupil Attributes and Engagement

Educational research has identified a number of pupil attributes that influence behavior in classrooms and learning. Since one attribute is unlikely to influence all types of engagement, a useful form of criterion-related validity in this context is to identify relationships between specified pupil attributes and engagement categories. For example, pupils with high formal reasoning ability may be more likely to volunteer explanations requiring higher cognitive level behavior than pupils with low formal reasoning ability. Conversely, a significant relationship between pupil engagement rates and formal reasoning ability may not be expected for lower cognitive level behaviors. Tobin and Capie (1981 a, b) reported statistically significant relationships between formal reasoning ability and two engagement modes : generalizing ($r = .23$, $p < .01$) and comprehending ($r = .19$, $p < .05$). Only comprehending was

significantly related to formal reasoning ability when total engagement (overt plus covert) was used as the criterion for analysis ($r = .23$, $p < .01$).

Locus of control, defined as the tendency for an individual to accept responsibility for academic success or failure is a variable that is likely to influence engagement rates. Pupils with an internal locus of control might be expected to engage to a greater extent in situations in which their efforts could lead to academic success. Tobin and Capie (1981 a, b) reported locus of control to be significantly related to pupil attending ($r = .21$, $p < .01$). When total engagement rates were considered, locus of control was significantly related to quantifying ($r = .22$, $p < .001$) and planning ($r = .20$, $p < .01$).

Significant correlations of formal reasoning ability and locus of control with specific engagement categories support the viewpoint that pupil engagement should be categorized. If global definitions such as on-task and off-task are used there is a likelihood that relationships with pupil attributes may not be discovered. As a consequence, when individuals are the focus of research, categorization of pupil engagement is advocated so that understandings can be developed of relationships between attributes of individuals and classroom behavior. For example, since pupil attending behavior was related to locus of control, knowledge of locus of control could help a teacher maintain higher engagement levels among pupils with an external locus of control. In a long term sense, strategies might be devised to

assist pupils in becoming more internal. The significant relationships between formal reasoning ability and rates of generalizing, and comprehending are interesting. Perhaps pupils with high levels of formal reasoning could formulate responses that were applicable to the generalizing and comprehending categories while other pupils could not. Another possibility is that teachers only called on pupils who were capable of providing the desired interpretation or justification.

Of course other pupil attributes are also strong contenders for consideration in process-product studies in classrooms. For example, pupil perceptions of the socio-psychological learning environment were reported by Ellett, Capie, and Johnson (1980) to be positively related to a measure of pupil on-task time.

Data Analysis

Because of the central position of pupil engagement in the process-product model for classroom learning, research hypotheses can be constructed with engagement as a dependent variable or an independent variable. Strong logical arguments can be advanced for causal links between the adjacent cells of the model in Figure 1, particularly when variables are manipulated. As a consequence, causal modeling is advocated involving the use of statistical analyses such as: path analysis; structural analyses using programs such as LISREL (Joreskog and Sorbom, 1978); and econometric analyses such as two- and three-stage least squares regression. Engagement data to be used in

such analyses would be proportions of the observed frequency of a particular engagement mode compared to the total number of observations.

Insert Figure 1 about here

When increased understanding of learning is the focus of research, individual pupil data often constitute the most logical and easily interpretable unit for analysis. For example, individual pupil engagement measures can be used as a basis for aptitude-treatment interaction studies. Individual pupil engagement measures can also be used along with pupil attributes in studies involving pupil achievement.

Individual data can also be aggregated to form a class mean to be used as the unit for analysis. Research questions related to validation of teacher performance measures (e.g. Capie, Ellett, and Johnson, 1980; Capie, Tobin, and Bowell, 1980) often use class means as validating criteria. One point that needs to be emphasized is that interpretation of results at the class level may differ substantially from interpretation of results based on individual pupil data. As is always the case, however, the appropriate level of aggregation of data is dependent on the research questions that provide a basis for the study.

Summary

Compelling reasons for measuring pupil engagement relate to the potential influence of the variable on pupil learning. Despite the intuitive appeal of the argument that pupils learn when they are directly involved in what is to be learned, additional research in a range of subjects is needed to substantiate this view. Research conducted so far suggests that the type of pupil engagement that is necessary for learning is dependent on the type of learning and attributes of the learner. In an endeavor to increase the effectiveness of classroom learning, research involving pupil engagement as an independent and a dependent variable needs to be initiated. Priority should be directed toward identification of achievement-related types of engagement and teaching strategies that can be used to promote desired classroom processes.

A fundamental concern of any research is for the dependability of the data used in hypothesis testing. Pupil engagement measures are no exception. Data collection procedures need to be adopted so that individual pupils or classes can be differentiated on the basis of the engagement data. G-studies can be planned to guide design of a data collection system. When research hypotheses have been tested, generalizability coefficients should be calculated for each type of engagement as a guide to interpretation of results.

As research on classroom learning continues, the number of variables included in a teaching-learning model is likely to increase,

and greater concern will focus on causal links between variables. Consequently, use should be made of data analysis procedures that can accommodate large numbers of variables and allow causal hypotheses to be tested.

Pupil engagement is indeed a fertile area for research, with numerous methodological concerns as well as substantive questions to be addressed as the complex interactions of the classroom are unravelled.

References

Capie, W., Ellett, C.D., and Johnson, C.E., Relating pupil achievement gains to secondary student teachers' performance. Paper presented at the annual meeting of the Eastern Educational Research Association, Norfolk, Virginia, March 1980.

Capie, W., Tobin, K.G., and Bowell, M., Using science achievement to validate student teacher competencies. A paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA, April, 1980. (ERIC Document Reproduction Service No. ED186261).

Carroll, J.H., A model of school learning, Teachers College Record, 1963, 64, 723-733.

Crandall, V.C., Katkovsky, W., & Crandall, V.J., Children's beliefs in their own control of reinforcements in intellectual -- academic situations, Child Development, 1965, 36, 91-109.

Hecht, L.W., Measuring student behavior during group interaction, Journal of Educational Research, 1978, 71, 283-290.

Joreskog, K.G. and Sorbom, D., LISREL IV - A general computer program for estimation of a linear structural equation system by maximum likelihood methods. Chicago: National Educational Resources, 1978.

Ladaherne, H.M., Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms, Journal of Educational Psychology, 1968, 59, 320-324.

Samuels, S.J., & Turnure, J.E., Attention and reading achievement in first-grade boys and girls, Journal of Educational Psychology, 1974, 66, 29-32.

Smith, P.L., Sampling Errors of Variance Component Estimates in Small Multifacet Generalizability Studies, Journal of Educational Statistics, 1978, 3, 319-346.

Smith, P.L., Some Approaches to Determining the Stability of Estimated Variance Components, A paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA, April 1980.

Stallings, J., Allocated academic learning time revisited, or beyond time on task, Educational Researcher, 1980, 9 (11), 11-16.

Tobin, K.G., and Capie, W., Relationships between formal reasoning ability, locus of control, academic engagement and integrated process skill achievement. Journal of Research in Science Teaching. 1981 a, 19, (in press).

Tobin, K.G., and Capie, W., Relationships between classroom process variables and middle school science achievement. Journal of Educational Psychology (submitted paper), 1981 b.

Tobin, K.G., and Capie, W., An empirical investigation of the stability of variance components and generalizability coefficients derived from teacher performance data. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April, 1981 c.

Table 1

Expressions for Estimating Variance Components in Three-Facet and Four-Facet Hierarchical Designs

Source of Variation	Number of Levels	Variance Component Estimate
(a) Three-facet design		
Subject S	S	$\frac{MS_S - MS_{L(S)}}{PL}$
Lessons L(S)	L	$\frac{MS_{L(S)} - MS_{RES}}{P}$
Sampling Occasion	P	
(b) Four-facet design		
Class C	C	$\frac{MSc - MS_{S(C)}}{PLS}$
Subjects S(C)	S	$\frac{MS_{S(C)} - MS_{L(S(C))}}{PL}$
Lessons L(S(C))	L	$\frac{MS_{L}(S(C)) - MS_{RES}}{P}$
Sampling Occasion	P	

Table 2

Variation in the Magnitude of the Generalizability Coefficients for Individual Pupil Enagement as the Number of Lesson and Observations within Lessons are Varied

Observation within Lessons	Lessons					
	1	2	3	4	5	6
1	.12	.22	.30	.36	.41	.46
2	.20	.33	.43	.50	.55	.60
3	.25	.40	.50	.57	.62	.67
4	.29	.45	.55	.62	.67	.71
5	.31	.48	.58	.65	.70	.73
10	.41	.58	.68	.73	.78	.81
15	.43	.60	.69	.75	.79	.82
20	.44	.62	.71	.76	.80	.83
25	.46	.63	.72	.77	.81	.83
30	.47	.64	.72	.78	.81	.84
35	.47	.64	.73	.78	.82	.84

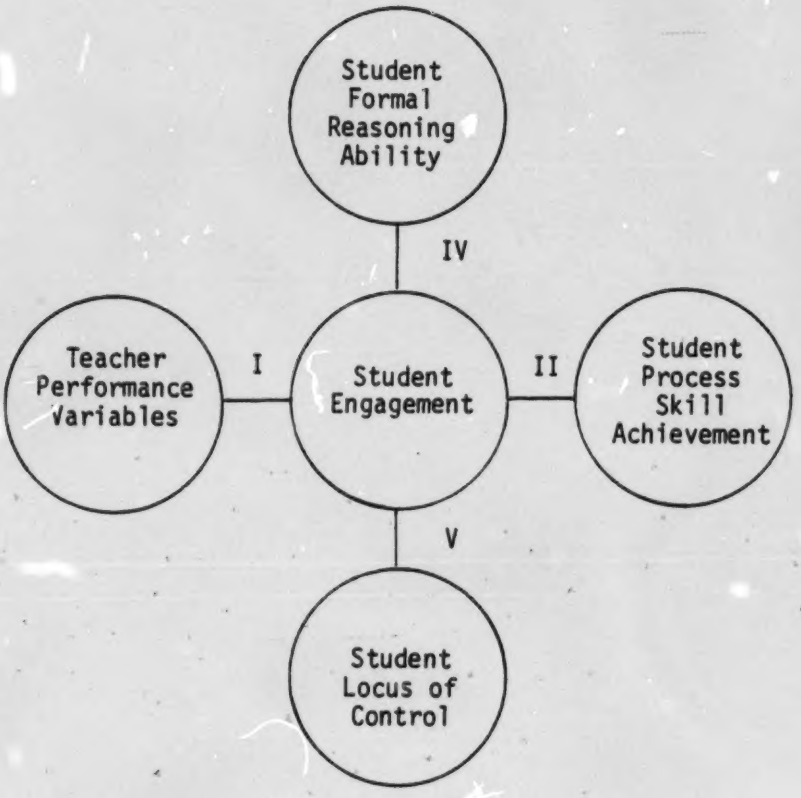

Figure 1. A process-product model for student achievement of science processes